exercises 3

Research Methodology – Quantitative

As part of an experiment investigating text entry using a stylus on soft keyboards, MacKenzie and Zhang (2001) evaluated if **entry speed** by *stylus-tapping* could be predicted from **user's speed** in *touch-typing* with a standard keyboard. The experiment involved 12 participants. The participants were given a pre-test to measure their touch-typing speeds. During the experiment, participants entered text using a stylus and a Qwerty soft keyboard displayed on an LCD tablet and digitizer. The pre-test touch-typing speed (independent variable) and experimentally measured stylus-tapping speed (dependent variable) are given in Figure 7.12a for each participant.

(a)

Participant	Stylus Tapping Speed (wpm)	Touch Typing Speed (wpm)
P1	18.2	42
P2	23.6	44
P3	26.0	32
P4	20.3	50
P5	20.3	36
P6	17.1	33
P7	24.0	74
P8	14.7	22
P9	20.3	31
P10	19.7	33
P11	22.4	25
P12	13.1	19

(b)

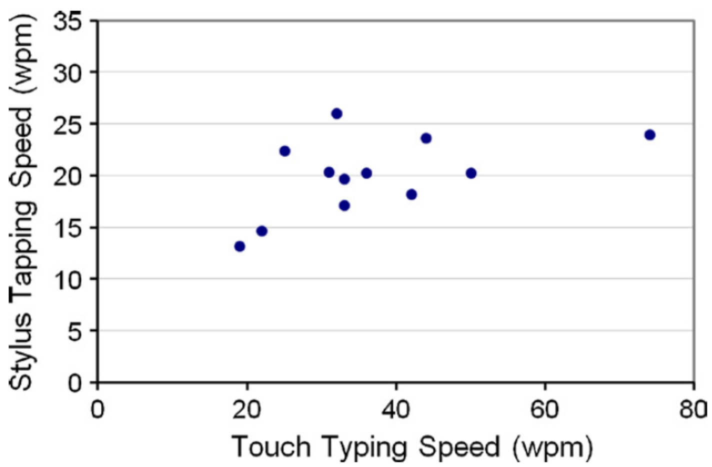

FIGURE 7.12

Relationship between stylus-tapping speed and touch-typing speed: (a) Data. (b) Scatter plot.

(Adapted from MacKenzie and Zhang, 2001, Fig. 4)

Exercise 3.1 Construct a data frame called D containing the variables reported in the table showed in Figure 7.12.

Exercise 3.2 Reproduce the graphical representation shown in Figure 7.12 according to all its details (use all necessary arguments of the graphical function to obtain a copy of this graphical representation).

Exercise 3.3 Run a regression model analysis by using TTS (Touch Typing Speed) as independent variable and STS (Stylus Tapping Speed) as dependent variable. Comment the result of the analysis and superimpose the estimated regression line on the previous graphical representation (Ex. 3.2).

Exercise 3.4 Extract the residuals of the regression model and plot them as a function of the independent variable. Moreover use the `abline()` function to represent the theoretical line corresponding to the abscissa (x-axis of the Cartesian plane). **Note** that if the residuals are correctly represented in the linear regression model, then the pair (x_i, e_i) should be randomly located around this line.

Exercise 3.5 The function `pt(tx,df)` in the R environment can be used to integrate the t-distribution with df degrees of freedom from $-\infty$ to the numerical value tx, that is to say, it computes the area under the curve (represented by a t-distribution with df degrees of freedom) from $-\infty$ to the numerical value tx. Use this function to compute the p-values for the beta0 parameter and beta parameter under the assumption that the null hypotheses are: H_0^1 (beta0 = 0), H_0^2 (beta = 1). Use the output of the linear regression analysis (see Ex. 3.4) to set the arguments of the `pt()` function. Comment the results and provide an interpretation for the null regression model defined by the two hypotheses H_0^1 and H_0^2.

Exercise 3.6 Run a new regression analysis this time on the standardized values of the two variables TTS and STS. Comment the results (including the inferential decisions) of the new analysis and provide a new graphical representation showing the relation between the two variables and the corresponding linear regression model. Verify that the sample correlation between TTS and STS corresponds to the new estimated slope in the linear regression. Finally, re-compute the p-values under the null hypotheses defined in Ex. 3.5. Comments the new inferential results.